

Converting Clinical Data into Information: Mapping Operational Hospital Data into a Time-Oriented Clinical Repository

Charles Rogerson, Ph.D.¹, Vinay Sabharwal², William W. Woo¹, John Foy, M.D.³

¹Systems Architecture Group, TDS Healthcare Systems Corporation, San Jose CA,

²VS Associates, Fremont CA, ³TDS Healthcare Systems Corporation, Atlanta GA

Email: crogerson@tds-hsc.com

ABSTRACT

The growing demand within healthcare enterprises for distributed, ad hoc query access to operational data integrated from multiple heterogeneous data sources can be met through time-oriented clinical repositories stored in SQL-based relational databases using client-server architecture. Correctly mapping, transforming and loading operational data from proprietary non-relational data structures into relational schemas in an automated way when such data structures may change over time is a complex problem that has not been generally solved. We present a general solution based on data modeling, schema transformation and configurable mapping dictionaries. Large quantities of data from the operational patient care information systems at a cancer research hospital have been successfully mapped and loaded into relational databases to support clinical research, CQI, outcomes studies, decision support and other population-oriented analysis. Today's healthcare environment, with constantly changing regulatory and reimbursement requirements and rapidly evolving delivery systems, places unprecedented pressures for information on the clinical and financial leadership of the healthcare enterprise (HCE). Personnel at many levels of the HCE must be able to access and analyze the data now residing in multiple systems and data sources with ease and flexibility. Most clinical data still resides in *paper chart form* as unstructured text and typically is also maintained in multiple electronic stores and ancillary systems. To allow for *ad hoc* access to the data for a large patient population requires a methodology for Information Mapping that supports the transformation of diverse data from multiple data sources into a time-oriented clinical information repository. The Information Mapping methodology is used for defining the source and target systems as well as the mapping between the multiple data stores. It begins with the creation of source and target conceptual schemas and results in the population of the meta-data dictionaries. The initial conceptual source schema represents the

initial scope of source data and the first cut of the source groups. It is important to note that the source data model is not a true, normalized data model. Rather, it is simply a collection of various bundles of data assembled into a collection of logical groups. The target conceptual schema represents the data that will be stored in the repository as a relational database. It clearly prescribes what must be captured and/or derived from the source systems. The target schema presently contains 21 tables and a total of 239 columns. The tables include Patient, Address History, Episode, Order, Physician, Diagnosis History, Pharmacy Charting, and Result, among others. These conceptual schemas are then embellished into logical schemas. The source and target logical schemas are the two key components upon which the information mapping is based. Once the source logical schema is completed two major tasks can begin. First the source schema will be used to map the data elements to the source logical groups. All the meta-data is captured into a separate data repository that catalogues all the relevant information. Once the target logical schema has been completed and validated, it must be defined in the meta-data repository. Once the source data elements (logical groups) along with the target (tables, columns) and the source and target schemas have been entered in the RML dictionaries, the last, but most important step is the information mapping itself. The mapping defines the data source and transformation for each field of a specific target schema. This mapping definition is also entered into the meta-data repository. The information mapping step relies on the accuracy of the data definitions of the source and target schemas and therefore is highly sensitive to changes in either the meta-data of each element or the source or target schemas. Very often modifications will be necessary in one or more of the logical schemas or the logical grouping; this will also affect the information mapping. It is recommended that a uniform procedure be instituted to regulate and synchronize such changes.